Speech Enhancement and Dereverberation With Diffusion-Based Generative Models

Year: 2022 | Citations: 304 | Authors: Julius Richter, Simon Welker, Jean-Marie Lemercier, Bunlong Lay, Timo Gerkmann

---

Abstract

In this work, we build upon our previous publication and use diffusion-based generative models for speech enhancement. We present a detailed overview of the diffusion process that is based on a stochastic differential equation and delve into an extensive theoretical examination of its implications. Opposed to usual conditional generation tasks, we do not start the reverse process from pure Gaussian noise but from a mixture of noisy speech and Gaussian noise. This matches our forward process which moves from clean speech to noisy speech by including a drift term. We show that this procedure enables using only 30 diffusion steps to generate high-quality clean speech estimates. By adapting the network architecture, we are able to significantly improve the speech enhancement performance, indicating that the network, rather than the formalism, was the main limitation of our original approach. In an extensive cross-dataset evaluation, we show that the improved method can compete with recent discriminative models and achieves better generalization when evaluating on a different corpus than used for training. We complement the results with an instrumental evaluation using real-world noisy recordings and a listening experiment, in which our proposed method is rated best. Examining different sampler configurations for solving the reverse process allows us to balance the performance and computational speed of the proposed method. Moreover, we show that the proposed method is also suitable for dereverberation and thus not limited to additive background noise removal.